

IMPLEMENTASI SMOTE UNTUK MENGATASI *CLASS IMBALANCE* PADA NAIVE BAYES *CLASSIFIER* DALAM ANALISIS SENTIMEN CALON PRESIDEN DI PEMILIHAN UMUM 2024

ABSTRAK

Susi Setianingsih

Euforia Pemilu 2024 sudah mulai terasa sejak tahun 2022, khususnya untuk pemilihan presiden dan wakil presiden. Banyak lembaga di Indonesia melakukan survei untuk memetakan kekuatan dan melihat peluang kandidat dalam berkontestasi. Akan tetapi, cara ini dinilai kurang efektif, sehingga perlu alternatif lain seperti akuisisi media sosial salah satunya Twitter atau 'X'. Analisis sentimen pada platform media sosial X dapat digunakan untuk memperoleh informasi dari berbagai macam sentimen oleh pengguna melalui *tweet*. Teknik Naive Bayes *Classifier* digunakan untuk klasifikasi sentimen karena memberikan hasil yang cukup baik. Namun, data *tweet* cenderung tidak seimbang pada setiap kelasnya, sehingga diperlukan suatu pendekatan untuk mengatasi ketidakseimbangan kelas yaitu dengan menggunakan SMOTE (*Synthetic Minority Oversampling Technique*). *Data set* diambil melalui API token dari aplikasi X. Data *tweet* akan di-*labelling* untuk selanjutnya dilakukan pemodelan. Hasil akhir dilakukan dengan menentukan model dengan nilai akurasi tertinggi. Untuk memudahkan dalam melakukan proses analisis, tampilan dalam bentuk website akan dipilih sebagai pendukung dalam melakukan analisis sentimen. Tampilan website akan dibuat menggunakan framework Streamlit dengan bahasa pemrograman Python. Hasil yang diperoleh menunjukkan bahwa Uji 3 dengan 2400 *data training* mendapatkan nilai akurasi terbaik mencapai 72.80%. Terdapat peningkatan nilai akurasi sebesar 15-23% lebih tinggi di kelompok uji yang menggunakan implementasi SMOTE pada model Naive Bayes. Perbedaan jumlah data yang digunakan juga memiliki pengaruh terhadap perbaikan hasil nilai akurasi pengujian sebesar 1-10%.

Kata kunci: Ketidakseimbangan Kelas, Naive Bayes *Classifier*, Pemilihan Umum, SMOTE, X

**IMPLEMENTATION OF SMOTE TO OVERCOME CLASS IMBALANCE IN
NAIVE BAYES CLASSIFIER IN SENTIMENT ANALYSIS OF
PRESIDENTIAL CANDIDATES IN 2024 GENERAL ELECTION**

ABSTRACT

Susi Setianingsih

The euphoria of the 2024 elections has begun to be felt since 2022, especially for the presidential and vice presidential elections. Many institutions in Indonesia conduct surveys to map the strength and see the chances of candidates in contesting. However, this method is considered less effective so that other alternatives are needed such as the acquisition of social media, one of which is Twitter or 'X'. Sentiment analysis on social media platform X can be used to obtain information from a wide variety of sentiments by users through tweets. The Naive Bayes Classifier technique is used for sentiment classification because it provides quite good results. However, tweet data tends to be unbalanced in each class, so an approach is needed to overcome class imbalance by using SMOTE (Synthetic Minority Oversampling Technique). The data set is retrieved through the token API from application X. The tweet data will be labeled for further modeling. The final result is done by determining the model with the highest accuracy value. To facilitate the analysis process, a website will be chosen as a support for sentiment analysis. The website display will be created using the Streamlit framework with the Python programming language. The results obtained show that Test 3 with 2400 training data gets the best accuracy value reaching 72.80%. There is an increase in accuracy value of 15-23% higher in the test group that uses the SMOTE implementation in the Naive Bayes model. The difference in the amount of data used also has an influence on improving the results of the test accuracy value by 1-10%.

Keywords: *Class Imbalance, General Election, Naive Bayes Classifier, SMOTE, X*